MODEL REGRESI ROBIT DAN PENERAPANNYA

Survinky, Erna Tri Herdiani, La Podje Talangko
Program Studi Statistika, Jurusan Matematika FMIPA Universitas Hasanuddin
Email: survinky7@gmail.com

Abstrak

Dalam analisis data kategorik, model probit dan logistik adalah model yang paling populer digunakan untuk memodelkan hubungan antara peubah respon yang kualitatif dengan peubah prediktor yang kuantitatif maupun kualitatif. Akan tetapi menurut beberapa ahli dalam tulisannya, model probit dan logistik tidaklah *robust* terhadap keberadaan data pencilan. Dalam tulisan ini, dibahas mengenai model robit yang merupakan model alternatif yang robust untuk model logistik dan model probit. Model ini menggunakan fungsi distribusi kumulatif dari distribusi t-student sebagai pengganti fungsi distribusi kumulatif dari distribusi normal pada model regresi probit. Adapun estimasi parameter regresi dari model regresi robit dilakukan dengan metode *Maximum Likelihood* melalui Algoritma EM. Model robit ini diterapkan pada data berbentuk (x_i, y_i), dimana y_i adalah variabel respon kategorik yang biner, dan x_i adalah variabel prediktor berbentuk vektor kovariat yang terdiri dari $x_{i0}, x_{i1}, x_{i2}, \dots, x_{ij}$. Hasil yang diperoleh adalah model peluang yaitu $P_r(y_i = 1)$ ataupun $P_r(y_i = 0)$.

Kata Kunci : analisis data kategorik, distribusi t-student, model regresi robit, taksiran maksimum likelihood, algoritma EM.

Abstract

In categorical data analysis, probit and logistic models are most popular model used to be modelling the relationship between the qualitative response variable with quantitative or qualitative predictor variables. But according to some experts in their paper, probit and logistic models are not robust to outliers. This paper discussed about the robit models that are robust alternative models of logistic and probit models. This model use the cumulative distribution function of the student-t distribution as a replacement the cumulative distribution function of the normal distribution on the probit regression model. As for the parameter estimates of the robit regression model performed by Maximum Likelihood method via EM-Algorithm. This robit model applied to the data which form (x_i, y_i), where y_i is the binary categorical response variables, and x_i is the predictor variabel which shaped on covariate vector, consisting of $x_{i0}, x_{i1}, x_{i2}, \dots, x_{ij}$. The results was obtained is a model of opportunity, that is $P_r(y_i = 1)$ or $P_r(y_i = 0)$.

Key Words : categorical data analysis, t-student distribution, robit regression model, maksimum likelihood estimation, the EM-algorithm.

I. PENDAHULUAN

Analisis regresi merupakan salah satu metode statistika untuk mendeskripsikan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Dalam menentukan model regresi, satu hal yang penting diperhatikan adalah struktur data dari variabel respon tanpa mengabaikan fungsi dan peranan dari variabel prediktor. Apabila variabel respon bersifat kualitatif maka metode mencari hubungan antara variabel respon Y terhadap prediktor X dapat menggunakan model logistik ataupun probit.

Pada umumnya model logistik dan probit digunakan untuk memodelkan hubungan antara variabel respon dikotomus atau biner dengan variabel prediktor kontinu ataupun kategori, tetapi dengan modifikasi logistik dan probit dapat digunakan ketika variabel respon bersifat politomus yaitu terdiri lebih dari 2 kategori. Adapun model logistik menggunakan *cumulative distribution function* (c.d.f) dari distribusi logistik sementara model probit menggunakan *cumulative distribution function* (c.d.f) dari distribusi normal standar. Akan tetapi, menurut Liu (2004) penduga Maksimum Likelihood untuk model probit adalah tidak *robust* atau tidak kekar terhadap *outlier*.

Beberapa ahli seperti Rubin (1983); Lange, Little, and Taylor (1989); Liu and Rubin (1995) dalam Liu (2004) telah mencoba menggunakan distribusi-t student dalam berbagai konteks dimana variabel respon biasanya dimodelkan dengan distribusi normal untuk mengestimasi parameter secara *robust*. Dengan mengganti distribusi normal pada regresi probit dengan distribusi-t student, maka diperoleh sebuah model yang kemudian disebut *regresi robit*. Penggunaan model ini sebagai alternatif dari regresi logistik dan probit sebelumnya juga disarankan oleh Mudholkar and George (1978) serta Albert and Chib (1993) dalam tulisannya

II. TINJAUAN PUSTAKA

2.1. Regresi Logistik dan Probit

Secara umum, model regresi dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Untuk kasus dengan peubah respon kualitatif (kategorik), persamaan regresi ditransformasi menjadi probabilitas. Menurut Kleinbaum dan Klein (2002), model regresi logistik, dimana variabel respon Y adalah biner (dua kategori), dapat didefenisikan sebagai berikut:

$$P(Y_i = 1 | X_1, X_2, \dots, X_k) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \quad , i = 1,2, \dots, k \qquad (1)$$

Model regresi probit merupakan hasil modifikasi dari model regresi logistik dengan menetapkan persamaan regresi logistik mengikuti distribusi normal. Model regresi probit secara umum dapat dituliskan sebagai berikut:

$$\begin{aligned} & P(Y_i = 1 | X_1, X_2, \dots, X_k) \\ & = 1 - P(Y_i = 0 | X_1, X_2, \dots, X_k) \\ & = \Phi\left(X_i' \beta\right) \qquad (2) \end{aligned}$$

Dimana $\Phi(X_i' \beta)$ adalah fungsi distribusi kumulatif (c.d.f) dari distribusi normal standar (Liu, 2004).

2.2. Model Robust

Suatu data dikatakan baik apabila data tersebut berada di sekitar garis regresi. Kenyataannya, terkadang terdapat data yang terletak jauh dari garis regresi atau pola data keseluruhan. Data tersebut dikenal dengan istilah pencilan atau *outlier*. Namun, *outlier* tidak dapat dibuang atau dihapus begitu saja dari pengamatan. Menurut Draper dan Smith (1992) dalam (Hasanah, 2012)., adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik data lainnya. Oleh karena itu, diperlukan suatu estimasi yang bersifat robust atau tahan terhadap pencilan yang dikenal dengan regresi *robust*.

Regresi robust diperkenalkan oleh Andrews (1972) dan merupakan metode regresi yang digunakan ketika distribusi dari *residual* tidak normal dan atau mengandung beberapa *outlier* yang berpengaruh pada model. Tujuan utama regresi *robust* adalah untuk memberikan hasil yang stabil karena kehadiran pencilan (Ryan, 1997).

2.3. Distribusi *t-student* Dalam *Robust Statistical*

Distribusi t-student dalam perkembangannya mempunyai peranan sebagai alat untuk inferensi *robust statistical* (Lange *et al.*, 1989; Liu and Rubin, 1995;). Kebanyakan pengunaan distribusi t-student dalam *robust procedure* ialah pada model statistik yang didasarkan pada distribusi normal yaitu dengan menggantikan distribusi normal dengan distribusi t-student pada model tersebut. Hal ini dikarenakan inferensi statistika yang didasarkan pada distribusi normal diketahui rentan terhadap *outlier* (Lange *et al.*, 1989).

Distribusi t-student dapat dituliskan $t_v(\mu, \sigma^2)$ dimana v adalah derajat kebebasan, μ adalah rataan, dan σ adalah skala parameter. Fungsi kepadatan peluang (*p.d.f*) dari distribusi t-student dituliskan sebagai berikut:

$$\boldsymbol{ft_v(x)} = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v \sigma^2}}\left(\mathbf{1} + \frac{1}{v}\frac{(x-\mu)^2}{\sigma^2}\right)^{-(v+1)/2} \qquad (3)$$

(Figueiredo, 2004)

III. PEMBAHASAN

3.1. Model Regresi Robit

Misalkan sebuah data pengamatan terdiri dari n pengamatan $X_{obs} = \{(x_i, y_i): i = 1, \dots, n\}$ dengan vektor kovariat x_i berdimensi p dan y_i variabel respon yang bernilai 0 atau 1. Model regresi biner yang *robust* untuk data tersebut dapat diperoleh dengan menggantikan distribusi normal pada model regresi probit dengan distribusi t-student. Model ini memberikan pendekatan yang *robust* untuk regresi logistik yang banyak digunakan dalam analisis data kualitatif (Liang *et al*, 2010). Model regresi biner dengan distribusi t-student ini kemudian disebut model regresi *robit* (Liu, 2004).

Secara umum model regresi robit dapat dituliskan sebagai berikut:

$$\boldsymbol{y_i Robit = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij}}$$

Dimana,

$$\boldsymbol{P_r(y_i = 1) = 1 - P_r(y_i = 0) = F_v(x_i^T \beta)} \qquad (4)$$

y_i adalah variabel respon biner dan x_i adalah vektor kovariat yang terdiri dari $x_{i0}, x_{i1}, x_{i2}, \dots, x_{ij}$, dimana $x_{i0} = 1$. β adalah parameter regresi yang tidak diketahui yang terdiri dari β_0, β_1, β_2, ..., β_j, sedangkan F_v adalah fungsi distribusi kumulatif dari distribusi t-student dengan rataan 0, skala parameter 1, dan derajat bebas v.

Untuk menaksir parameter dari model regresi *robit*, dapat dilakukan dengan *Maximum Likelihood Estimation* (MLE) ataupun dengan Estimasi Bayesian.

3.1.1. Estimasi Parameter Model Regresi Robit dengan metode Maksimum Likelihood

Diketahui $X_{obs} = (x_i, y_i)$, y_i variabel respon kategorik berdistribusi Bernaulli dimana bernilai 0 dan 1, sedangkan x_i adalah vektor kovariat terdiri dari $x_{i0}, x_{i1}, x_{i2}, \dots, x_{ij}$, dimana $x_{i0} = 1$. Diketahui fungsi kepadatan peluang dari distribusi bernaulli adalah $f(y|p) = p^y(1-p)^{1-y}$, maka fungsi kepadatan peluang model robit adalah sebagai berikut:

Untuk $i = 1$ maka

$$f\big((x_1, y_1)|\beta\big) = \big[F_v(x_1'\beta)\big]^{y_1}\big[1 - F_v(x_1'\beta)\big]^{1-y_1}$$

Untuk $i = 2$ maka

$$f\big((x_2, y_2)|\beta\big) = \big[F_v(x_2'\beta)\big]^{y_2}\big[1 - F_v(x_2'\beta)\big]^{1-y_2}$$

…

…

…

Untuk $i = n$ maka

$$f\big((x_n, y_n)|\beta\big) = \big[F_v(x_n'\beta)\big]^{y_n}\big[1 - F_v(x_n'\beta)\big]^{1-y_n}$$

Sehingga diperoleh persamaan likelihood untuk model robit sebagai berikut:

$$f((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n); \beta) = f(X_{obs}; \beta)$$
$$= f\big((x_1, y_1)|\beta\big) f\big((x_2, y_2)|\beta\big) \dots f\big((x_n, y_n)|\beta\big)$$

$$f(X_{obs}; \beta) = \prod_{i=1}^{n} f\Big(\big(x_i, y_i\big)|\beta\Big)$$

$$\boldsymbol{f(X_{obs}; \beta) = \prod_{i=1}^{n}[F_v(x_i'\beta)]^{y_i}[1 - F_v(x_i'\beta)]^{1-y_i} \quad (\beta \in \mathbb{R}^p)} \quad (5)$$

Persamaan likelihood di atas akan digunakan untuk mengestimasi paremeter β. Akan tetapi menurut *Lewandowsky et al* (2010), β tidak bisa diperoleh secara langsung dari persamaan MLE diatas. Untuk mendapatkan MLE dari β,dapat digunakan algoritma dan berbagai type dari algoritma EM itu sendiri seperti ECME dan PX-EM (Liu, 2004).

3.1.2. Complete-Data Untuk Penerapan Algoritma EM

Diketahui algoritma EM adalah algoritma yang digunakan untuk mendapatkan taksiran ML dari parameter model yang dalam penerapannya senantiasa melibatkan variabel laten. Variabel laten sendiri adalah variabel yang secara tidak langsung teramati. Artinya, mungkin saja ada nilai-nilai yang hilang dalam data, atau dengan kata lain model dapat dirumuskan dengan mengasumsikan adanya tambahan titik data yang tidak teramati.

Dalam penerapan algoritma EM untuk mengestimasi parameter (β) dari model regresi robit, model *complete-data* dispesifikasikan dengan mengasumsikan adanya data hilang berupa variabel laten (τ_i, z_i), untuk $i = 1,2, \dots, n$ dimana:

$$\tau_i|\beta \sim Gamma(v/2\,, v/2\,), \qquad (6)$$

$$z_i|(\tau_i, \beta) \sim N\big(x_i'\beta, 1/\tau_i\big) \qquad (7)$$

sehingga,

$$y_i = \begin{cases} 1, jika\ z_i > 0 \\ 0, jika\ z_i \le 0 \end{cases} \quad i = 1,2, \dots n \qquad (8)$$

Model *complete-data* tersebut termasuk dalam keluarga eksponensial dengan statistik cukup untuk β:

$$S_{\tau x x'} = \sum_{i=1}^{n} \tau_i x_i x_i' \quad dan \quad S_{\tau x z} = \sum_{i=1}^{n} \tau_i x_i z_i \qquad (9)$$

dan model taksiran *Maksimum Likelihood* untuk β diberikan sebagai berikut:

$$\hat{\beta}_{com} = S_{\tau x x'}^{-1} S_{\tau x z} \qquad (10)$$

3.1.3. Algoritma EM untuk Estimasi Parameter Model Robit

Untuk mengestimasi parameter β dari model robit dengan menggunakan algoritma EM, *E-step* dan *M-step* dilakukan dengan cara sebagai berikut:

* *E-step*
  Hitung seluruh τ_i dan z_i kemudian hitung ekspektasi statistik cukup untuk β yaitu:
  $E\left(S_{\tau x x'} \middle| \beta = \beta^{(t)}\right) = \hat{S}_{\tau x x'}$ dan $E\left(S_{\tau x z} \middle| \beta = \beta^{(t)}\right) = \hat{S}_{\tau x z}$
* *M-step*
  Dapatkan taksiran β yang baru yaitu $\beta^{(t+1)} = \hat{S}_{\tau x x'}^{-1} \hat{S}_{\tau x z}$

Lakukan *E-step* dan *M-step* secara iteratif dimulai dari $\beta^{(0)} = (0, \dots, 0)$ dengan *t* menunjukkan jumlah iterasi hingga diperoleh selisih nilai-nilai $\beta^{(t+1)}$dan $\beta^{(t)}$semuanya mendekati 0 atau dengan kata lain kurang atau sama dengan δ (nilai yang sangat kecil mendekati nol).

Misalkan diketahui $f_v(x)$ adalah fungsi kepadatan peluang dari $F_v(x)$ maka τ_i dan z_i dalam *E-step* dapat dituliskan dengan menggunakan hasil yang diperoleh dalam Liu (2004) sebagai berikut:

* $\hat{\tau}_i = E\left(\tau_i \middle| \beta = \beta^{(t)}, X_{obs}\right)$

$$= \frac{y_i - (2y_i - 1)F_{v+2}(-(1+2/v)^{\frac{1}{2}} x_i' \beta^{(t)})}{y_i - (2y_i - 1)F_v(-x_i' \beta^{(t)})} \qquad (11)$$

* $\widehat{\tau_i z_i} = E\left(\tau_i z_i \middle| \beta = \beta^{(t)}, X_{obs}\right) = \hat{\tau}_i \hat{z}_i \qquad (12)$
  Dimana:

$$\hat{z}_i \equiv x_i' \beta^{(t)} + \frac{(2y_i - 1) f_v(x_i' \beta^{(t)})}{y_i - (2y_i - 1)F_{v+2}(-(1+2/v)^{\frac{1}{2}} x_i' \beta^{(t)})} \qquad (13)$$

3.2. Penerapan Model Regresi Robit

Penerapan model regresi robit dalam penelitian ini dilakukan melalui simulasi dengan studi kasus menghitung peluang kekambuhan penyakit gastritis. Data yang digunakan adalah data sekunder berupa data hasil penelitian oleh Lenny Gannika yang digunakan dalam penyusunan sebuah karya ilmiah berjudul *"Hubungan Kebiasaan Sehari-Hari dan Stress Dengan Terjadinya Kekambuhan Gastritis Pada Pasien yang Dirawat di Puskesmas Batua Makassar" tahun 2009* dengan mengambil sebanyak 20 sampel secara berurut dari total 60 sampel dan 4 variabel dari total 11 variabel yang diamati dalam penelitian tersebut. Adapun data tersebut dapat dilihat dalam **Tabel 1** berikut:

Tabel 1. Data Kekambuhan Gastritis Pada Pasien di Puskesmas Batua Makassar tahun 2009

No	Nama	Umur (tahun)	Frekuensi makan	Kebiasaan merokok	Kekambuhan Gatritis
1	Ratna	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
2	Arna	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
3	Ahmad	10-40	Tidak teratur	Biasa merokok	Tidak kambuh
4	Dahari	>40	Teratur	Tidak merokok/tidak biasa	Kambuh
5	Thalasia	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
6	Harlina	>40	Teratur	Tidak merokok/tidak biasa	Tidak kambuh
7	Suhartini	10-40	Teratur	Tidak merokok/tidak biasa	Kambuh
8	Eni	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
9	Ruth	>40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
10	Dedi	10-40	Teratur	Biasa merokok	Kambuh
11	Jumiati	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
12	Rahel	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
13	Desi Nelly Kaya	10-40	Teratur	Tidak merokok/tidak biasa	Kambuh
14	Hamna	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
15	Nurhidayanti	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
16	Suharlini	10-40	Teratur	Tidak merokok/tidak biasa	Tidak kambuh
17	Aldi	10-40	Tidak teratur	Biasa merokok	Kambuh
18	Najamiah	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
19	Saleha	10-40	Tidak teratur	Tidak merokok/tidak biasa	Kambuh
20	Makmur	>40	Tidak teratur	Biasa merokok	Kambuh

Sumber: Data Penelitian Lenny Gannika tahun 2009

Berdasarkan data dari **Tabel 1**, diketahui bahwa $X_{obs} = (x_i, y_i)$, dimana y_i adalah kekambuhan gastritis, yang bernilai 0 untuk kambuh dan 1 untuk tidak kambuh. Sementara variabel x_i terdiri dari $x_{i0}, x_{i1}, x_{i2}, x_{i3}$, dimana $x_{i0} = 1$ atau $x_i = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix}$, dimana:

1. x_{i1} = umur, yang bernilai 0 untuk yang berumur lebih dari 40 tahun

dan bernilai 1 untuk yang berumur 10-40 tahun

2. x_{i2} = frekuensi makan, yang bernilai 0 untuk tidak teratur dan bernilai 1 untuk yang teratur
3. x_{i3} = kebiasaan merokok, yang bernilai 0 untuk tidak/ tidak biasa merokok dan 1 untuk yang biasa merokok

Model regresi robit dari data tersebut dapat dituliskan sebagai berikut:

$$P_r(y_i = 1) = 1 - P_r(y_i = 0) = F_v(x_i^T \beta)$$

dimana :

$P_r(y_i = 1)$ = Peluang Kekambuhan Gastritis
$P_r(y_i = 0)$ = Peluang Ketidakkambuhan Gastritis

β = koefisien regresi = $\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{bmatrix}$

F_v = Fungsi kumulatif distribusi t-student

β adalah parameter yang belum diketahui dalam model regresi robit ini. Oleh karena itu β akan diestimasi dengan menggunakan metode *Maksimum Likelihood* dengan cara menerapkan algoritma EM

Dengan bantuan software maple dan Microsoft Office Excel 2007 untuk menghitung dan menjalankan algoritma EM, diperoleh hasil hingga iterasi ke- 3

$$\beta^{(4)} = \begin{bmatrix}1.2269\\ 0.1726\\ -0.8346\\ -0.4884\end{bmatrix}.$$

Dengan demikian, misalkan estimasi parameter β dari model berhenti pada iterasi $t = 3$, maka diperoleh model regresi robit sebagai berikut:

$$\boldsymbol{y_i Robit = 1.2269 + 0.1726 x_{i1} + (-0.8346) x_{i2} + (-0.4884) x_{i3}}$$

dimana,

$$P_r(y_i = 1) = 1 - P_r(y_i = 0) = F_v(x_i^T \beta)$$

dan F_v adalah fungsi kumulatif dari distribusi t-student dengan rataan 0, skala parameter 1, dan derajat bebas $v = 19$.

Dari model tersebut, misalkan kita ingin menghitung seberapa besar peluang kekambuhan gastritis pada seorang perokok berumur lebih dari 40 tahun yang frekuensi makannya tidak teratur, maka:

$$P_r(y_i = 1) = F_v(x_i^T \beta)$$

$$= \int_{-\infty}^{x_i'\beta} f_v\left(x_i'\beta\right) d(x_i'\beta)$$

$$= \int_{-\infty}^{x_i'\beta} \frac{\Gamma\left(\frac{v+1}{2}\right)}{(\pi v)^{\frac{1}{2}} \Gamma\left(\frac{v}{2}\right)\left(1+\frac{(x_i'\beta)^2}{(v)}\right)^{\frac{v+1}{2}}} d(x_i'\beta)$$

$$= \int_{-\infty}^{0.738} \frac{\Gamma(10)}{(19\pi)^{\frac{1}{2}} \Gamma(9.5)\left(1+\frac{(x_i'\beta)^2}{(19)}\right)^{10}} d(x_i'\beta)$$

$$= \frac{\Gamma(10)}{(19\pi)^{\frac{1}{2}} \Gamma(9.5)} \int_{-\infty}^{0.738} \frac{1}{\left(1+\frac{(x_i'\beta)^2}{(19)}\right)^{10}} d\left(x_i'\beta\right)$$

$$= \frac{362880}{7.72(119292.46)}(1.94350769)$$

$$= 0.765806$$

Jadi, peluang kekambuhan gastritis pada seorang perokok berumur lebih dari 40 tahun yang frekuensi makannya tidak teratur adalah sebesar 0.765806 atau 76.58%.

VI. KESIMPULAN

Model regresi robit merupakan salah satu model regresi yang digunakan untuk menganalisis data dengan variabel respon yang bersifat kualitatif (kategorik). Taksiran *maksimum likelihood* parameter β untuk model regresi robit dengan algoritma *Expectation Maximization* (EM) adalah:

$$\beta^{(t+1)} = \hat{S}_{\tau x x}^{-1}{}'\hat{S}_{\tau x z}$$

dimana proses penaksiran terhadap parameter β dilakukan secara iteratif pada dua tahap algoritma yaitu *E-Step dan M-Step.*

Model regresi robit adalah model peluang dan dapat diterapkan pada data yang berbentuk (x_i, y_i) yang diketahui derajat bebasnya, dimana y_i adalah peubah respon kategorik, seperti pada data kekambuhan gastritis di puskesmas batua tahun 2009. Model peluang kekambuhan Gastritis pada seseorang berdasarkan data yang digunakan setelah iterasi ketiga pada estimasi parameternya adalah:

$$\boldsymbol{y_i Robit = 1.2269 + 0.1726 x_{i1} + (-0.8346) x_{i2} + (-0.4884) x_{i3}}$$

Dari model yang diperoleh dapat diketahui bahwa peluang kekambuhan gastritis pada seorang perokok berumur lebih dari 40 tahun yang frekuensi makannya tidak teratur adalah sebesar 0.765806 atau 76.58%.

DAFTAR PUSTAKA

Agresi, A. (2010). *Historical Highlights in the Development of Categorical Data Analysis*. Department of Statistics, University of Florida.

Albert, J.H., Chib, S. (1993). *Bayesian Analysis of Binary and Polychotomus Response Data.* Journal of the American Statistics Association. **88**:422, 669-679.

Figuiredo, M.A.T., (2004). *Lecture Notes on The EM Algorithm.* Instituto de Telecomunicacoes, Instituto Superio Tecnico. Portugal.

Gannika, L., (2009). *Hubungan Kebiasaan Sehari-Hari dan Stress Dengan Terjadinya Kekambuhan Gastritis Pada Pasien yang Dirawat di Puskesmas Batua Makassar*. Program Studi Ilmu Keperawatan Fakultas Kedokteran. Universitas Hasanuddin, Makassar.

Gujarati, D. 1997. *Ekonometrika Dasar*. Dra. Ak. Sumarno Zain, MBA, penerjemah. Jakarta: Erlangga. Terjemahan dari: *Basic Econometrics.*

Hasanah, I. (2012). *Regresi Robust Untuk Mengatasi Outlier Pada Regresi Linier Berganda.* Prodi Matematika Fakultas Sains dan Teknik, Universitas Jendral Soedirman, Purwokerto.

Ilma, A. (2009). *Taksiran Persamaan Likelihood Pada Model Persamaan Struktural Non Linear*. Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

Kleinbaum, D.G., Klein, M. (2002). *Logistic regression: a self-learning text (second edition)*. Springer-Verlag, New York.

Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989), *Robust Statistical Modeling Using the t Distribution*, Journal of the American Statistical Association, 84, 881-896.

Lewandowski, A., Liu, C., & Vander Wiel, S. (2010). *Parameter Ekspansion and Efficient Inference.* Statistical Sciences. **25**:4, 533-544. Institute of Mathematical Statistics.

Liang, S., Liu, C., & Carrol, R.J. (2010). *Advanced Markov Chain Monte Carlo Methods: Learning From Past Sample.* Willey, London.

Liu, C. (2004). *Robit Regression: a Simple Robust Alternative to Logistic and Probit*. Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives (A. Gelman and X. L. Meng, *eds.*) 227-238. Willey, London.

Liu, C and Rubin, D. B. (1995). *ML Estimation of the Multivariate t Distribution With Unknown Degrees of Freedom*. Statistica Sinica, **5**, 19-39.

Mudholkar, G.S. and George E.O. (1978). *A remark on the shape of the logistic distribution*. Biometrika, **65**, 667-668.

Ryan, T.P. (1997). *Modern Regression Methods*. New York : A Wiley-Interscience Publication.

Saumi, T.F. (2013). *Analisis Regresei Terapan: Regresi Probit.* Departemen Statistika Fakultas MIPA, Institut Pertanian Bogor (IPB).

Widiharih, T. (2009). *Buku Ajar Statistika Matematika II.* Universitas Diponegoro, Semarang.

Yong, B. (2003). *Penaksir Maksimum Likelihood Bagi Model Probit dan Model Probit Bivariat*. Integral majalah ilmiah MIPA, **8:**1.